

This is a repository copy of *How should we do the history of big data?*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/97690/>

Version: Accepted Version

Article:

Beer, David Gareth orcid.org/0000-0002-6926-4595 (2016) How should we do the history of big data? Big Data & Society.

<https://doi.org/10.1177/2053951716646135>

Reuse

["licenses_typename_other" not defined]

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

This is the accepted version of an article to be published in *Big Data & Society*.

How should we do the history of big data?

David Beer

Abstract

Taking its lead from Ian Hacking's article 'How should we do the history of statistics?', this article reflects on how we might develop a sociologically informed history of big data. It argues that within the history of social statistics we have a relatively well developed history of the material phenomenon of big data. Yet this article argues that we now need to take the concept of 'big data' seriously, there is a pressing need to explore the type of work that is being done by that concept. The article suggests a programme for work that explores the emergence of the concept of big data so as to track the institutional, organisational, political and everyday adoption of this term. It argues that the term big data has the effect of making-up data and, as such, is powerful in framing our understanding of those data and the possibilities that they afford.

How should we do the history of big data?

Around 25 years ago, in a piece that was revised for publication in the highly influential collection *The Foucault Effect*, Ian Hacking (1991) asked the question ‘how should we do the history of statistics?’. An apparently straightforward question that is likely to provoke some complex answers. What I would like to do here is to revisit that question in light of the emergence of ‘Big Data’¹. Put simply, I’d like to ask the question: how should we do the history of big data? Again this might seem straightforward, but by asking this question I am hinting at two things. First is the argument that we need to contextualise our understandings of big data within the history of social statistics. That is to say that we need to place big data within the genealogy of social data of various types. Second is the argument that we should approach this history by treating big data as both a material phenomenon and also a concept. Indeed, my central argument here is that we need to explore the *concept of big data* in historical, political and sociological terms. This is important because ‘big data’ is a concept that has achieved a profile and vitality that very few concepts attain. As such, its influence needs to be unpicked and understood. We need to understand the work that is being done by this powerful and prevalent concept.

To get things started though, what I am proposing here is that when thinking about big data we need to consider its history as being tied-up with particular ways of thinking. We then need to consider how this thinking is enacted in the development of certain infrastructures and in the industry of data analytics. This is to see big data as the entwinement of both a phenomenon and a concept. Big data itself, with its capacity to track lives through archived and classified forms of individuated data, can be placed then within the genealogical lineage of the modern state (Beer, 2016). In this sense, it could be argued that we already have a history of big data that can be found in accounts of the history of the use of statistics to know and govern populations (see for example Desrosières, 1998; Porter 1986 & 1995; Hacking 1990; Mackenzie, 1981; and Foucault, 2007; which is discussed in Elden, 2007). Similarly, we also have sociological resources that enable us to understand the power dynamics that reside within these accumulating data about people and populations (instructive examples here are

Espeland & Sauder, 2007; Espeland & Stevens, 2008; and the various essays collected in Rottenberg et al, 2015). We even have a discussion of the use of big data in historical work – which asks how big data can be used by historians to gather or archive resources (Manning, 2013). And yet closer to the aims of this particular article, Halpern (2014), who focuses predominantly on the aesthetics of data, provides a historical account of the emergence of data since 1945. Despite all of this though we have something that is yet to be explored from a historical and sociological perspective, which is the work that is done by the very concept of ‘big data’. That is to say that we have little understanding of the concept itself, where it came from, how it is used, what it is used for, how it lends authority, validates, justifies, and makes promises. In other words, we now need to work through a detailed account of what might be thought of as *the birth of big data*.

To reiterate, this stream of work is not concerned with the data itself, but with the discourse, terminology and rhetoric that surrounds it and which ushers and affords its incorporation into the social world. This is not to say that the specific material properties of big data are somehow unimportant, but rather that the way that these data are framed in particular rationalising discourses is also needs to be treated carefully if we are to form a more detailed appreciation of the social implications of those data. It could be argued that in many ways the power dynamics of big data are to be found just as much in the way that those data are labelled and described as it is in the actual data themselves. Indeed, given the difficulties of data access and the technical and computing skills required to analyse big data, it might even be argued that the concept has far greater reach than the material phenomenon. However far the material consequences of big data might reach, the rationalities within through which it is understood are likely to reach further. This article is dedicated to beginning to open-up this stream of work and is geared to developing a more contextual account of the concept of big data as it becomes embedded in organisational, political, social, cultural and everyday life.

The ‘avalanche’ of numbers

In Hacking’s (1991) aforementioned essay on the history of statistics, which links into his other more substantial works on the same topic, he discusses some of the features of what he refers to as the ‘avalanche of numbers’. This is where we can start to begin to contextualise the so-

called big data revolution within a much longer history. Given the date of his piece it is obvious that when referring to this avalanche of numbers Hacking is not talking about the rise of digital technologies, smartphones, wearables or social media. Indeed, he is actually talking about an 'avalanche of numbers' that occurred around 1820 to 1840 as a new 'enthusiasm for numbers' (Hacking, 1991: 186) and a growing assemblage for data gathering took hold. Elsewhere this same period has been referred to as experiencing an 'explosion' of numbers. As Porter (1986: 11) observes, the 'great explosion of numbers that made the term statistics indispensable occurred during the 1820s and 1830s'. This 'explosion' or 'avalanche' of data occurred as 'nation-states classified, counted and tabulated their subjects anew' (Hacking, 1990: 2). In other words, the sense that we are being faced with a deluge of data about people is not something that is entirely new, in fact it has a long history. The type of data may have changed as might its analytics – with the shift toward commercial and algorithmic forms amongst other changes – but the lineage is clear. There are, of course, features of the current data moment that are in some ways novel but it is still interesting to note that this idea of a scaling up of social data, the feeling that we are facing an unfathomable flow of social data, itself has a history. The notion of an 'avalanche' gives the sense of the weight of escalating data resources that is comparable with the notion that data have suddenly got *big*. Both are based upon the feeling that there is a sudden and unstoppable wash of flowing data about people, a pooling of data that is on a scale that was not previously imagined. We have then both the phenomenon of the data interweaving with the way that it is imagined – the key difference here is that Hacking's terminology is based upon an *observation* about the increasing role of metrics in the 19th Century whereas big data is a term that is commonly used in everyday discourse to refer to the data phenomenon of that very moment. Big data is a concept which, like the data and methods associated with it, has a 'social life' (Savage, 2013). Referring back to the 19th Century, Hacking (1991: 189) concludes that 'almost no domain of human enquiry is left untouched by the events that I call the avalanche of numbers, the erosion of determinism and the taming of chance'. We see here that already there was a sense that data harvesting was spreading out across the social world, and that this was accompanied by new means for analysing patterns in that data and for dealing with questions of probability.

By way of illustration let us turn again to Hacking's essay on how we should do the history of statistics. In that piece Hacking (1991: 191) points out that the 'the avalanche of numbers is at least in part the result of industrialization and the influx of people from the country to the town'. It was with the move towards industrialisation and the centralization of large parts of national populations in urban environments that the statistics about people began to escalate. The possibilities that came with these infrastructural changes were accompanied, Hacking argues, by a 'sheer fetishism for numbers' (Hacking, 1991: 192) and cultural shifts associated with the 'new countings' or 'new numberings' (Hacking, 1991: 191). This combination of social, technological and cultural changes led to an expansion of data that Hacking refers to as the 'avalanche of numbers'. As well as being counted in new ways, populations were also then ordered through categorisations. As Hacking (1991: 192) explains, 'when the avalanche of numbers began, classifications multiplied because this was the form of this new kind of discourse'. The emergence of new metrics also led to people being classified in new ways, which had powerful implications for how individuals and groups were perceived and treated. This new type of social ordering emerged with the need to manage the accumulating data about people. Hacking's (1991: 182) point here is that 'many of the modern categories by which we think about people and their activities were put in place by an attempt to collect numerical data'. Thus new categories emerged through which these new data might be gathered and by which they might then be analysed – leading to all kinds of 'classificatory struggles' (Tyler, 2015).

Part of the power associated with these escalating numbers, Hacking proposes, was to be found in their apparent objectivity. Statistical data, Hacking (1991: 184) claims, 'have a certain superficial neutrality'. It is this very appearance of neutrality that lends them an air of authority and which makes them so powerful. Tied in with this neutrality is the ability to use the numbers and categories to define what is seen to be normal and what is therefore seen to be abnormal. As Hacking (1991: 183) puts it, 'there are also statistical meta-concepts of which the most notable is 'normalcy''. Thus these accumulating data became a central means by which populations could be known and governed, and where understandings and expectations were produced alongside powerfully reinforced norms. Put simply, Hacking's (1991: 183) observation is that 'statistics of populations...form an integral part of the

industrial state'. Industrial modernity brought with it expanded archives of data about populations (see Featherstone, 2000).

The result of all this is that statistics have become an important component in governance. Hacking's argument is that norms and classifications based around these types of data enable social facts to be brought into existence. Hacking (1991: 181) argues that:

'Statistics has helped determine the form of laws about society and the character of social facts. It has engendered concepts and classifications within the human sciences. Moreover the collection of statistics has created, at the least, a great bureaucratic machinery. It may think of itself as providing only information, but it is itself part of the technology of power in a modern state'.

Statistics are then incorporated into the very infrastructures and modes of governance of the state – in the last twenty or so years we might also add corporate and commercial data gathering to this. The result is that the categories and modes of reasoning surrounding these data become part of the formal and legal structures of the state, with direct implications for how people are treated (for one example in relation to immigration see Schinkel, 2013). Again, in Hacking's (1991: 194) words, the 'bureaucracy of statistics imposes not just by creating administrative rulings but by determining classifications within which people must think of themselves and of the actions that are open to them'. These emerging numbers quickly came to define how people saw themselves, how they saw others and, complimenting these, how limits and boundaries were placed around actions and opportunity. So, big data can be placed within this long history of social statistics, but we might also note that there is undoubtedly an intensification in the scale of data over that time, particularly as commercial organisations have joined in with the state to increase the infrastructures, scope, accumulation and deployment of data (Kitchin, 2014; Beer, 2016; Ajana, 2013).

As this would suggest it is important to situate this article within the history of the development of statistics but this is beginning to take us into territory that resides beyond the remit of this article, what I'd like us to take from this is that the expansion of data and metrics is not something that can be isolated to a particular moment in the recent past. Rather the powerful ordering presence of data has been felt for some time, as has the sense that there is an overwhelming deluge of information associated with the march of modernity.

What I would like to do here is to put this particular historical context to one side, there are other places we might go in order to explore the genealogical history of what is now referred to as big data (for an overview of the history of social statistics in relation to big data see chapter 2 in Beer, 2016). Instead, let us be aware of the infrastructural, technical and cultural history of such data whilst focusing our attention upon the relatively short life of the actual concept or term *big data*. We may need further work that delimits the particular material and ontological properties of the current form that this social data takes – such as the work being conducted by Kitchin (2014; see also Kitchin & McArdle, 2016) – but let us turn our attention elsewhere for the moment. The concept of big data has a short history which is part of a much longer series of developments stretching back hundreds of years. I want to use the remainder of this article though to argue that it is the work that is being done by this particular concept that requires our attention – particular if we are to continue to attempt to develop a more complete and contextual understanding of the influence of data today. As such, the point is that the history of big data as a phenomenon can be tracked back through the pages of those histories of social statistics – even if much more work is needed in order for a more global and ‘connected’ (Bhambra, 2014) history of statistics to be developed – but what we have little understanding of is the *birth* and *life* of big data as a concept. It is this concept that needs to be tracked, unpacked and examined. Indeed, the aim of this article is to begin to map-out a programme of work that needs to be completed in order for us to fully understand the politics of big data.

Treating ‘big data’ as a concept

This is where I would like us to prize ourselves away from the data themselves, to begin to think historically about how these data are conceptualised. It is by acknowledging the long history of the accumulation of data about individuals and populations that we can begin to make a departure into seeing the different ways that data are presented in conceptual terms – and thus where we might begin to see more clearly the importance of the project of exploring big data as an interweaving of a material phenomenon and circulating concept.

Both Ian Hacking and Stuart Elden suggest that the only way to really understand the power and influence of concepts is to see them in their historical context. Hacking’s (1991: 184)

position is that we need to explore 'the relationship between concepts in their historical site' (Hacking, 1991: 184). Similarly, Elden (2013a: 15) argues that 'conceptual history is important because of its emphasis on terminology, and the relation between meaning and designation; contextualist approaches are crucial in stressing the importance of reading texts within the situations in which they were written'. In relation to territory Elden's (2013a: 15) position is that 'territory is a word, concept and practice; and the relations between these can only be grasped historically' – this is a project that Elden (2013b) expands upon in much greater detail in his book *The Birth of Territory*. The point here is that we can only understand certain social phenomena through their discursive and conceptual formulations, and we can only understand these conceptual formulations by thinking historically about them. Both Hacking and Elden place concepts at the centre of their historical analyses.

Hacking offers further explanation of his position by claiming that:

'the organization of our concepts, and the philosophical difficulties that arise from them, sometimes have to do with their historical origins. When there is a radical transformation of ideas, whether by evolution or by an abrupt mutation, I think that whatever made the transformation possible leaves its mark upon subsequent reasoning.' (Hacking, 1991: 184)

Concepts are a product of their historical origins, we might conclude, but they then also have social reach and influence themselves. The organization of our concepts can then be at the heart of social transformations – the transformation of ideas is a powerful thing. These concepts and the transformations of which they are a part leave, Hacking suggests, an indelible mark on future reasoning. They leave their mark on the way that the social world is comprehended and acted upon. If we are to pursue the concept of big data with this in mind, then we would not just be looking at the concept for its influence during the lifetime of its use but also its potential influence on future reasoning. We would also need to look at the discursive frameworks and modes of reasoning that fed into the concept of big data. Thus a genealogy of a concept like big data aims to capture the emergence of a concept as a part of a historical lineage of reasoning that shoots out into the past and the future. It is a moment, but a moment in which we might reveal something longer term.

We can of course see the influence of Michel Foucault echoing through Hacking and Elden's approaches. We can amplify these echoes by turning to a relatively well-known interview with Foucault which was originally published in 1980. The interview focuses upon questions of method. Amongst various aspects of Foucault's approach discussed in that interview, a particular theme emerges concerning the role of concepts in shaping social realities. Here Foucault describes some of the methods he deployed in his works and focuses in particular upon the need to explore conceptual processes in the formation of the social world. He focuses upon his concern with understanding the different ways in which truth is produced through practice. As Foucault explains:

'To put the matter clearly: my problem is to see how men govern (themselves and others) by the production of truth (I repeat once again that by production of truth I mean not the production of true utterances, but the establishment of domains in which the practice of true and false can be made at once ordered and pertinent).'

(Foucault, 1991: 79)

It is unusual to find something so crucial hidden within brackets. Foucault is interested in exploring the ways in which truth is produced so as to see how those truths limit understandings, actions and practices. His intention is to use events and moments to open up these regimes of truth, and to understand how these regimes of truth activate practices. As he put it, 'eventualizing singular ensembles of practices, so as to make them graspable as different regimes of 'jurisdiction' and 'veridiction': that, to put it in exceedingly barbarous terms, is what I would like to do' (Foucault, 1991: 79). His intention then was to grasp the practices that translate into the boundaries and limits of jurisdictions and, alongside this, to see how truth is verified in different ways – it is notable that he argues elsewhere that markets are the 'sites of veridiction' (see Foucault, 2008:32). In doing this his aim is to 'resituate the production of true and false at the heart of historical analysis and political critique' (Foucault, 1991: 79). Elsewhere Foucault (2014: 7) describes this production or manifestation of certain regimes of truth as a process of 'alethurgy' – which is concerned with understanding the 'manifestation of truth' as central to the formation of power structures. These regimes of truth and their limited powers can then be seen to be found in the discourse surrounding certain practices.

At this point in the interview Foucault's attention shifts to the notion of 'programmes' in order to exemplify and explain these wider objectives. It could be read that when he talks of programmes he is talking about the set of practices in which regimes of truth are imagined and then made possible. He talks here of programmes of activity that are not always realised, but which can be used to explore how ideas are projected onto the social world. In the interview Foucault is questioned on the separation of these programmes from the reality of what is happening on the ground. Foucault's response is to emphasize the importance of understanding how the world is imagined in order to understand how it unfolds. As he explains:

'Bentham's *Panopticon* isn't a very good description of 'real life' in nineteenth-century prisons. To this I would reply: if I had wanted to describe 'real life' in the prisons, I wouldn't indeed have gone to Bentham. But the fact that this real life isn't the same thing as theoreticians' schemas doesn't entail that these schemas are therefore utopian, imaginary, etc. One could only think that if one had a very impoverished notion of the real. For one thing, the elaboration of these schemas corresponds to a whole series of diverse practices and strategies' (Foucault, 1991: 81)

Obviously referring back to the work he did for his 1975 book *Discipline and Punish*, Foucault is arguing that the types of programmes or imagined possibilities captured in concepts like the panopticon are important. These types of concepts become woven into reality in different ways, they become part of practice as they are cemented into jurisdictions, boundaries and as they verify, authorise and select what comes to be. Concepts, or programmes, can then be elaborated in practice in ways that are not always obvious. Separating them from reality would be a mistake. Thus we cannot see big data as being a programme that exists outside of the practices of the use of data on the ground. Similarly, big data may not necessarily be a very good concept for seeing the 'reality' of everyday life, but it is a good concept for understanding how visions of contemporary data are incorporated into the imagining of life, the production of truths and the liminal work that contains the social world. Big data is undoubtedly a part of contemporary strategies and practices. The point here is that big data may be treated as a programme of thought that needs to be analysed in this way.

Foucault extends this point further. He adds that these 'programmes induce a whole series of effects in the real (which isn't of course the same as saying that they take the place of the

real): they crystallize into institutions, they inform individual behaviour, they act as grids for the perception and evaluation of things' (Foucault, 1991: 81). As such, imagined programmes and conceptual formations translate into regimes of truth. That is to say that they solidify into practices, organisations, institutions and behaviours. To apply this to big data we could imagine how this concept carries with it 'grids for the perception and evaluation of things'. That is to say, that it is not just the evaluations that come from the applications of big data themselves, but that the concept of big data as a programmatic mode of reasoning also brings with it the values and norms that provide the means for evaluating and judging. It is not just the data that afford judgments, it is also the very concept of big data itself that shapes decisions, judgments and notions of value – as it brings with it a vision for particular types of calculative or numerical knowing about individuals, groups and the social world. These are legitimised in the case of big data by notions of its scale and the eradication of error and inefficiency (for a discussion of scale and accuracy in big data see boyd & Crawford, 2012). These programmes of big data arrive with a thirsty desire to render measurable.

For Foucault, whether or not these imagined programmes are ever fully realised is not necessarily important, rather it is the influence that those imagined programmes have in shaping practice. It is also the broader rationality that they reflect. In the case of big data we might not see the project or its imagined potential realised in full, but the concept has already been influential far beyond the reach of the data in many respects. Just because programmes are never fully realised does not mean that they are somehow insignificant, especially when they achieve the prominence that is enjoyed by the big data movement. Rather we should see how that programme was pursued and how the imagined outcomes became part of practices and strategies – or how they relate or encapsulate a broader art of governance, political economy or prevalent forms of rationality and reasoning. According to Foucault:

'These programmings of behaviour, these regimes of jurisdiction and veridiction aren't abortive schemas for the creation of reality. They are fragments of reality which induce such particular effects in the real as the distinction between true and false implicit in the ways men 'direct', 'govern', and 'conduct' themselves and others' (Foucault, 1991: 82)

Here the *programme* shifts to being about *programming*, about setting up the codes of social life. With these conceptual framing being fragments of reality. Such programmes fracture into

reality, for Foucault. It is in the job of unpicking these fragments that he is interested. Foucault's use of true and false might seem blunt, but he is pointing to the powerful ways in which such programmes set rigid limits. Such conceptual programmes, for Foucault then, 'induce effects' and make things happen. They are part of governance and they act to shape conduct by contributing towards these regimes of truth that inform behaviour. As he further explains, we need to attend to 'the correlative formation of domains and objects and...the verifiable, falsifiable discourses that bear on them; and it's not just their formation that interests me, but the effects in the real to which they are linked' (Foucault, 1991: 85). The challenge, once we take such schema as being important to the conduct of the reality of the social world, is in thinking about how to explore their emergence and effects.

Framing big data

Recently we have seen some attempts that begin to think through or suggest the need to think through the role of the concepts and discourses that surround data today. For example, Rob Kitchin has suggested that we need to look at the political and economic framing of big data. He suggests that we should look at 'how a powerful set of rationalities is being developed to support the roll-out and adoption of big data technologies and solutions' (Kitchin, 2014: 126). Kitchin indicates that this is part of a broader project whilst focusing his discussion across four 'major tasks': 'governing people', 'managing organisations', 'leveraging value' and 'producing capital'.

For Kitchin these underpinning rationalities need to be explored because they play such a potent part in the integration of big data. These rationalities are to be found in the discursive regimes of big data, and thus these regimes need detailed and careful attention in order to understand the power dynamics of big data. One way into this is to look at the logic that is woven into the big data movement. As Kitchin (2014: 126) puts it:

'The power of the discursive regimes being constructed is illustrated by considering the counter-arguments – it is difficult to contend that being less insightful and wise, productive, competitive, efficient, effective, sustainable, secure, safe, and so on, is a desirable situation. If big data provide all of these benefits, the regime contends that it makes little sense not to pursue the development of big data systems.'

Considering the connotations and implications of those powerful underpinning rationalities reveals the potency of the discourse here. Big data brings with it a force to comply and a rationality that is hard to critique or resist. This can potentially be seen to have a kind of neoliberal reasoning or rationality at its core, one based upon the use of data as the mechanism by which the model of the market may be rolled out across the social world (for a discussion see Beer, 2016). As a result of these considerations Kitchin concludes that ‘what is presently required, through specific case studies is a much more detailed mapping out and deconstruction of the unfolding discursive regimes being constructed.’ (Kitchin, 2014: 126). There are undoubtedly parallels here between Kitchin’s suggestion and the project I’m mapping out in this article. Kitchin gives us a starting point through which we might channel the type of observations drawn out from Foucault, Elden and Hacking’s work. It is pressing, as Kitchin has put it, that ‘given the utility of the data, there is a critical need to engage with them from a philosophical and conceptual point of view.’ (Kitchin, 2014: 185). What we need then are the conceptual and historical resources that will enable us to develop a richer understanding of the discourses and rationalities of big data. It is this point that we have reached, we need now to work out ways of expanding such a set of insights and to flesh out this approach. My suggestion is that we focus our attention centrally upon the term of big data itself and begin to explore it historically and conceptually. This will provide a focal point for responding to Kitchin’s more general call. Part of this will require us to not just challenge or dismiss but to carefully unpick ‘boosterist discourses declaring their positive disruptive effects.’ (Kitchin, 2014: 192). It is this unfolding discursive regime that needs attention – by illuminating the rhetoric orbiting around the concept of big data. There is undoubtedly more to big data than its discursive framing, it has material properties that make it big data (see Kitchin & McArdle, 2016), yet the particularities of that discursive framing shape those material presences and the integration of big data into broader social structures and orders.

Elsewhere there are some other rare occasions in which the power of such discourses have been acknowledged. For library information scientist Ronald E. Day this discourse centres around a particular set of claims. The shift for Day is away from notions of ‘information’, which implies something flexible and informed, and toward something much more rigid. According to Day (2014: 3), ‘more recently, the discourse of “data”, conceived as a form of auto-affective presence or “fact,” has come to supersede the trope of “information”’. The shift then is

towards the notion that data is equivalent to facts, and thus then away from a more open vision of information. This, for Day, is an important shift that makes contemporary notions of data much more powerful in social formations. He continues this line of argument by claiming that these:

‘claims for knowledge are presented as immediate – “factual” – rather than as emergent through technologies, techniques, and methods, on the one hand, and interpreted through theory or a priori concepts, on the other hand. *The data says...; the data shows us...; we are only interested in data (not justifications/excuses/your opinion/your experience)...*; *big data and its mining and visualizations gives us a macroscopic view to see the world anew now* – these and similar phrases and tropes now fill the air with what is claimed to be a new form of knowledge and a new tool for governance that are superior to all others, past and present.’ (Day, 2014: 134; italics in the original)

This presentation of data as facts is crucial, for Day, in understanding the powerful role played by those data. Again, as with Kitchin, Day explores how the data are presented in compelling and even irresistible ways. In the above passage Day offers some illustrations of how this type of discursive framing works in practice. In these formulations the data is seen to be objective, neutral and telling – it is not something to be questioned or interrogated it is rather a social fact around which behaviour should be bent. It is seen to be a tool for governance that cannot be questioned or rivalled with subjective opinions. Data is seen, in this formation, to be unquestionable, accurate and over-arching in its panoramic view of the social world.

These positions provide some revealing opening insights, but we have not really gone much further than this acknowledgement that there is a need to think about the conceptual and discursive frames that accompany these data. It is this project that needs to be attended to, with some urgency. This now needs sustained attention to build upon some of these insights and to reinvigorate the type of project that Foucault, if you will pardon the assumption, may have taken on were he to have been around to observe the emergence or birth of big data.

If we were to explore the history of big data in terms of the history of the concept then that would lead us to try to understand the work that this concept does to shape practices and behaviours, to limit jurisdictions and to establish truths and desired outcomes. In short, it is

to explore the world views or perspectives that the term big data is woven from and provokes. This would be to approach the term big data as being built in the tensions of veridiction, and to see how it authorises certain behaviours, actions and outcomes. To see what perspectives and notions of truth that it endorses. To see how it brings with it a set of preferences and desires that it then legitimises. This is to see how the term big data itself has political ends as it comes to demarcate value or worth. We can certainly start by thinking about how the concept evokes certain feelings of trust through its apparent properties of objectivity and neutrality. In short, *the concept of big data frames and makes-up the data themselves*. With this in mind we need to see what type of work it does, how it leads us to see those data and how this framing is woven with particular ways of seeing that social world. The framing of the data is particularly powerful in this regard and will dictate not only what we get from the data but also the possibilities that are afforded simply from uttering these two words together.

Big data's 'grids of perception': focal points for analysing the making-up and framing of big data	Analytical questions and issues
Promises	What promises are being made in the discussion of the data? What hopes and futures are evoked or imagined? How is big data seen to promise possible outcomes, efficiencies, improvements and forms of progress?
Manifestations of truth	How is the data to be used to distinguish truth from falsity? What possibilities are presented in the truths implicit in the discussion of the data? What are the truths that big data is perceived to enable us to reveal, discover or uncover?
Jurisdiction formation and maintenance	How is big data used to set the territories of knowledge? How are boundaries placed around what can be known? Who is responsible for deciding what can be known through the data? How is this position policed and controlled? Who decides what is knowledge and who has the right to use and know it?
Veridiction	How is big data seen to present opportunities to verify, authorise and render appropriate? How is big data seen to legitimise and justify? What is then seen to be afforded by these legitimising processes? What do these systems verify and why? How does this link to truth making?
The demarcation of value & worth	How is big data used to frame what is seen to be valuable or worthwhile? How is big data used in the boundary work required to demarcate value? How is big data used to promote certain forms of value and to devalue other things? What are the implicit values laced into discussions of big data? Can worth be measured?
Limits placed on practice and behaviour	How is big data used to justify and present preferred practices and behaviours? How are the limits drawn around the acceptability of behaviours? How is big data used to define normality and abnormality? How is the sharing of practices implicated with certain limits that are woven into the discussion of big data?

Objectivity and neutrality	<p>How is big data presented as being an objective form of knowledge?</p> <p>To what extremes is this form of objectivity taken?</p> <p>How is the data presented as providing the basis of neutral forms of decision making or decision making at a distance?</p> <p>What are the questions of agency raised by the concept of big data and how is the responsibility for decision making shifted to these data?</p>
Judgments and evaluations	<p>What types of judgments and evaluations is big data seen to make possible?</p> <p>How is big data seen to afford processes of judgment and evaluation?</p> <p>What judgments and evaluations are being made and with what types of temporality, frequency and strength of outcome?</p> <p>What opportunities are presented for challenging those judgments? Or is big data seen to present the means for incontestable judgment and evaluation?</p>

Table 1: An analytic framework for structuring the analysis of the work done by the concept 'big data'.

This approach will require us to look across different sectors to see how big data, the term, has been used. It will look at how it has been deployed in commercial, political, economic and organisational discourse, and what type of work it has done in these sectors. Focusing upon this will hopefully then open up broader political motifs as they find their way into the language of everyday governance and social ordering.

Based on these discussions, Table 1 attempts to summarise the key analytical points that will be required to extend this project. Table 1 provides an analytical framework for exploring the work that is being done by the concept of big data. The left hand column presents the analytical focal points and the right hand column presents the types of questions and issues raised by those focal points. These are intended as analytical points of departure that will reveal the implicit dynamics of the concept of big data and will thus let us analyse it as a programme of activity and a way of thinking that becomes realised in the limits and practices of the social world. This framework is a heuristic that can be used to guide and shape our analysis, but it may well need to be adapted. The suggestion then is that the framework offered in Table 1 may be used to explore how the concept of big data is enacted and performed in making-up data across different social spheres and sectors. In short, this is an analytical framework that may be drawn upon wherever talk of big data may be found.

Conclusion

To conclude, I'd like to suggest that we have a relatively pronounced understanding of how data 'make people up', to use Hacking's (1990: 3) term, but we have relatively little appreciation of how *concepts make-up those data*. This is not to say that the particular material properties of the current data moment are unimportant, they clearly need further work to understand how those properties relate or differentiate them from that broader history of social statistics. It is to say though that what is also needed is a detailed exploration of the trajectory and influence of the concept of big data. We need to ask what work this term does, and what work it has done. We need to explore how it has become established within the discourse of organisations, funding bodies, political and policy circles, in journalism, in social commentary, and in various others sectors. We need to look at the emergence of this powerful concept and to understand how it has been shaped and reshaped in its use. We also need to understand how the term big data brings data to life, how it breathes life into data, how it makes them vital and telling.

Underpinning this approach is the pursuit of a more detailed understanding of how big data, as a concept, recrafts notions of value and worth. The concept of big data might seem unimportant – it might be dismissed as “business” or “managerial” talk, it might be seen as a passing fad, it might be seen to be part of the meaningless verbiage of contemporary media cultures – but the scale of the use of the term would suggest something different. The term big data is doing a lot of work, it is a persuasive presence in funding, management, decision making, “human capital” and the everyday practices of production and consumption. The work that is being done by the concept of big data needs attention, particularly as it is frequently doing far more than the actual data itself. Indeed, the term big data can be used to reveal the type of thinking and the mode of reasoning that ushers data and metric-led processes into everyday, organisational and social life. A part of the role that this concept plays concerns the different ways that it demarcates what is valued or what is seen to be worthwhile. It is a term that lends confidence, authority and objectivity to decisions that are then realised through the data themselves. This gives this particular term a very powerful social presence that needs to be unpicked. The threads then need to be followed back through the history of its usage.

All of this will require us to understand the visions within which and through which notions of big data are communicated. To see the way that big data is evoked and the kind of outcomes and sensibilities that it provokes. The power of big data is not just in the data themselves, it is in how those data and their potential is imagined and envisioned. To understand the power, influence and reach of big data requires us to understand the performative influence of the material data whilst also being attentive to the concept that frames them. My suggestion is that so far we have focused virtually all of our attention on the phenomenon and we have given very little attention to the powerful concept that defines, enacts and ushers in those apparently big data.

By looking back at some important historical accounts we can quickly see that what is most novel about big data is not necessarily the vast accumulation of data, although that is an important part or moment of an established and long set of genealogical threads, but the way that this concept of big data has taken on such commercial, organisational and economic force and power. For this reason, amongst others, I would suggest that we now need to lavish some attention on this loaded and powerful concept, particularly as it comes to define contemporary life in so many ways. It has been argued that when thinking of how our lives are measured we need to think about the modes or styles of thought that accompany that measuring, rather than just focusing upon the technical infrastructures (see Elden, 2006: 139-148; Hacking, 1990; Porter, 1986 & 1995). This is certainly true for big data. The pursuit of big data, like the pursuit of statistical measures of populations, is as much about a mode of reasoning or a way of thinking as it is about the assemblage that it generates.

The way to explore these modes of thought or styles of reasoning is to unpick and illuminate the role played by that very label of big data in various social spheres. Of course, any clear separation of the material phenomenon from the concept of big data is misleading, they work together and are intimately intertwined. My point here is that we need to think about the historical context in which big data is unfolding, we need to see it as part of the long series of developments in the measurement of people and populations. At the same time though, in pursuing a more contextual understanding we should not continue to be preoccupied with the data itself we also need to examine the type of data-thinking that is encapsulated in the term big data and in the use of that term. There is something to be said for this particular

moment in the long unfolding of metric based approaches to the social world, there are likely to be a number of things that are materially distinct about this particular moment in that history, but the things that need to be said require us to understand these apparently new forms of data whilst also paying careful attention to the way that they are packed, presented and rolled-out in the discourse that surrounds and permeates them. This may or may not be a unique or important moment in the history of social statistics and metrics, but it is nevertheless a moment in which a particular concept is taking hold and in which its power is worth some reflection. We simply cannot understand big data in historically informed and critical terms unless we analyse the interconnections between its materiality and the concept through which these material transformations are understood. What is perhaps most interesting about this moment in this long history is that we have such a prevalent and prominent term that presents this phenomenon to us as if it were a sudden and unique moment within that history.

Notes

1. In this article I will not offer a direct definition of 'big data' as such. This is for two reasons. First, this type of definition has been provided elsewhere, such as in Rob Kitchin's (2014) excellent and authoritative overview of big data (which includes a chapter detailing the definition of this term, see Kitchin 2014: 67-79). And, second, the approach that I outline in this article aims to explore the meanings and rationalities associated with the term 'big data'. As such, it aims to explore the various definitions that are attached to this particular term rather than treating it as a fixed entity. This article actually aims to use the term big data as a way into these types of defining statements and understandings – meaning that tying down the meaning too tightly from the outset may hamper its progress and scope.

References

Ajana, B. (2013) *Governing Through Biometrics: The Biopolitics of Identity*. Basingstoke: Palgrave Macmillan.

Beer, D. (2016) *Metric Power*. London: Palgrave Macmillan.

Bhambra, G.K. (2014) *Connected Sociologies*. London: Bloomsbury.

boyd, d. & Crawford, K. (2012) 'Critical questions for big data: provocations for a cultural, technological and scholarly phenomenon', *Information, Communication & Society* 15(5): 662-679.

Day, R. (2014) *Indexing it all: The subject in the age of documentation, information, and data*. Cambridge, Massachusetts: MIT Press.

Desrosières, A. (1998) *The Politics of Numbers: A History of Statistical Reasoning*. Cambridge, Massachusetts: Harvard University Press.

Elden, S. (2006) *Speaking Against Number: Heidegger, Language and the Politics of Calculation*. Edinburgh: Edinburgh University Press.

Elden, S. (2007) 'Governmentality, calculation, territory', *Environment and Planning D: Society and Space* 25(3): 562-580.

Elden, S. (2013a) 'How should we do the history of territory?', *Territory, Politics, Governance* 1(1): 5-20.

Elden, S. (2013b) *The Birth of Territory*. Chicago: Chicago University Press.

Espeland, W.N. & Sauder, M. (2007) 'Rankings and reactivity: How public measures recreate social worlds', *American Journal of Sociology* 113(1): 1-40.

Espeland, W.N. & Stevens, M.L. (2008) 'A Sociology of Quantification', *European Journal of Sociology* 49(3): 401-436.

Featherstone, M. (2000) 'Archiving cultures', *British Journal of Sociology* 51(1): 168-184.

Foucault, M. (1991a) 'Questions of Method', in, Burchill, G., Gordon, C. & Miller, P. (eds) *The Foucault Effect*. Chicago: The University of Chicago Press. pp. 73-86.

Foucault, M. (2007) *Security, Territory, Population: Lectures at the Collège de France 1977-1978*. Basingstoke: Palgrave Macmillan.

Foucault, M. (2008) *The Birth of Biopolitics: Lectures at the Collège de France 1978-1979*. Basingstoke: Palgrave Macmillan.

Foucault, M. (2014) *On the Government of the Living: Lectures at the Collège de France 1979-1980*. Basingstoke: Palgrave Macmillan.

Hacking, I. (1990) *The Taming of Chance*. Cambridge: Cambridge University Press.

Hacking, I. (1991) 'How should we do the history of statistics?', in, Burchill, G., Gordon, C. & Miller, P. (eds) *The Foucault Effect*. Chicago: The University of Chicago Press. pp.181-195.

Halpern, O. (2014) *Beautiful Data: A history of vision and reason since 1945*. Durham: Duke University Press.

Kitchin, R. (2014) *The data revolution: Big data, open data, data infrastructures & their consequences*. London: Sage.

Kitchin, R. & McArdle, G. (2016) 'What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets', *Big Data & Society* 3(January-June 2016): 1-10.

MacKenzie, D. (1981) *Statistics in Britain: The Social Construction of Scientific Knowledge*. Edinburgh: Edinburgh University Press.

Manning, P. (2013) *Big Data in History*. Basingstoke: Palgrave Macmillan.

Porter, T.M. (1986) *The Rise of Statistical Thinking 1820-1900*. Princeton: Princeton University Press.

Porter, T.M. (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.

Rottenburg, R., Merry, S.E., Park, S.J. & Mugler, J. (eds) (2015) *The World of Indicators: The Making of Governmental Knowledge Through Quantification*. Cambridge: Cambridge University Press.

Savage, M. (2013) 'The 'Social Life of Methods': A Critical Introduction', *Theory, Culture & Society* 30(4): 3-21.

Schinkel, W. (2013) 'The imagination of 'society' in measurements of immigrant integration', *Ethnic and Racial Studies* 36(7): 1142-1161.

Tyler, I. (2015) 'Classificatory struggles: Class, culture and inequality in neoliberal times', *The Sociological Review* 63(2): 493-511.